International Journal of Pure and Applied Mathematics
Volume 82 No. 1 2013, 143-161
ISSN: 1311-8080 (printed version)
url: http://www.ijpam.eu

HARMONIC ANALYSIS AND COINTEGRATION OF DIAGNOSTIC PARAMETER TIME SERIES OF MOTORS

Jean-François Emmenegger[1 §], Elena Pervukhina[2], Victoria Golikova[3]

[1]Department of Quantitative Economics
University of Fribourg
SWITZERLAND

[2,3]Sevastopol National Technical University
Universitetskaya Str., 33, 99053, Sevastopol, UKRAINE

Abstract: This study deals with time series of two parameters of *carburetor internal combustion engines* measured during industrial testing. The parameters are the *developed power* and the *fuel consumption rate per time unit*. The measurements are realised for both parameters at constant time interval of 0.4 seconds for the same period. First, for the time series evolving similarly over the period, a functional relation between the series is investigated through cointegration analysis, following the methodology of Johansen. Second, as the data show strong periodicity, an appropriated harmonic analysis is performed.

Key Words: 62M15, 62M10, 62P30, 91B84

1. Introduction

A petrol-powered carburetor four-cycle engine with 4 cylinders is investigated. The main goal of the study is to obtain authentic data on the latent defects of the motor working processes. The aim is to anticipate the arising of defects on the base of empirical information analysis long before their external display appear, see Rybalko [28]. But the specificity of the bench testing of motors

Received: December 7, 2012

§Correspondence author

(on one hand, complexity and non stationarity at deficient *a priori* information and, on other hand, restriction of material, labor, and available time) excludes the possibility to use well-known existing modeling and diagnostic methods to increase the testing efficiency. It is necessary to create a new approach to model engines with the goal to estimate their technical state on the basis of measured (diagnostic) parameters, see Skurikhin [30] and Klyuev [24].

In the scientific literature there are not numerous publications on the question. Moreover, the existing work is basically connected with traditional technologies of data processing to find correlation and regression of diagnostic parameters, therefore the results have no unequivocal physical interpretation.

New statistical methods, see Engle [5] and Johansen [16], have been developed specially to analyse economic time series. They allow to reveal the statistical dependence between non stationary random processes. The idea of the present analysis is to extend these techniques to processes of physical nature, describing the behaviour of tested objects in real time. These methods, usually used to model economic processes, open new opportunities to analyse experimental technical information, that are got by ordinary measuring devices during tests. The advantage of the method is simplicity and unambiguity of physical interpretation of the results, and also presentation and efficiency of the control. Moreover, they do not need additional measuring devices.

The aggregation of two scalar processes of fluctuating diagnostic parameters into one multivariate process is elaborated. For a *carburetor internal combustion engine* the main characteristics are:

a) *high-speed characteristics*, indicating the dependence of power, torque on the rotating velocity (rotation speed) of the cranked shaft, translating the linear piston motion into rotation,

b) *loading characteristics*, giving the dependent consumed fuel on power, torque or effective pressure of the engine at constant rotation speed of the crankshaft and

c) *adjusting characteristics*, which reflect dependence of developed power, torque and fuel consumption on parameter determining the state of the engine. For motors these parameters reflect the technical state of the motor.

Then, there are other technical characteristics like the roughness of surfaces, the deviation of sizes from nominal values, deviations of forms of real surfaces from nominal values, balancing of rotating parts and directly characterizing the quality of assemblage and adjustment of aggregates of the whole tested engine, see Golikova [11]. The roughness of details, size deviations from nominal values, from deviations of real surfaces from nominal forms are so-called *direct* or *structural* parameters. The developed power, torque and consumption of fuel are so-called *diagnostic* parameters.

The *time* to run motors, necessary to reduce the surface roughness, to correct macro-geometry is limited. The physical peculiarities of the working processes of motors, the measures of parameter processes at the stage of industrial testing are non stationary even in steady regimes.

This paper continues the approach to study the physically interpreted dependence between engine diagnostic parameters, modeled for complex problems. It was shown, see Golikova [11], [12] and Pervukhina [27], how they can be determined through multivariate stochastic analysis.

The methodology applied to analyse the time series is as follows. A *cointegration analysis* is performed, following the methodology of Johansen [16], establishing a linear cointegration relation between both time series. Then, the *power spectral density* (PSD) is estimated to get the variance of the noise and essential frequencies of the time series, see Priestley ([26], p. 395).

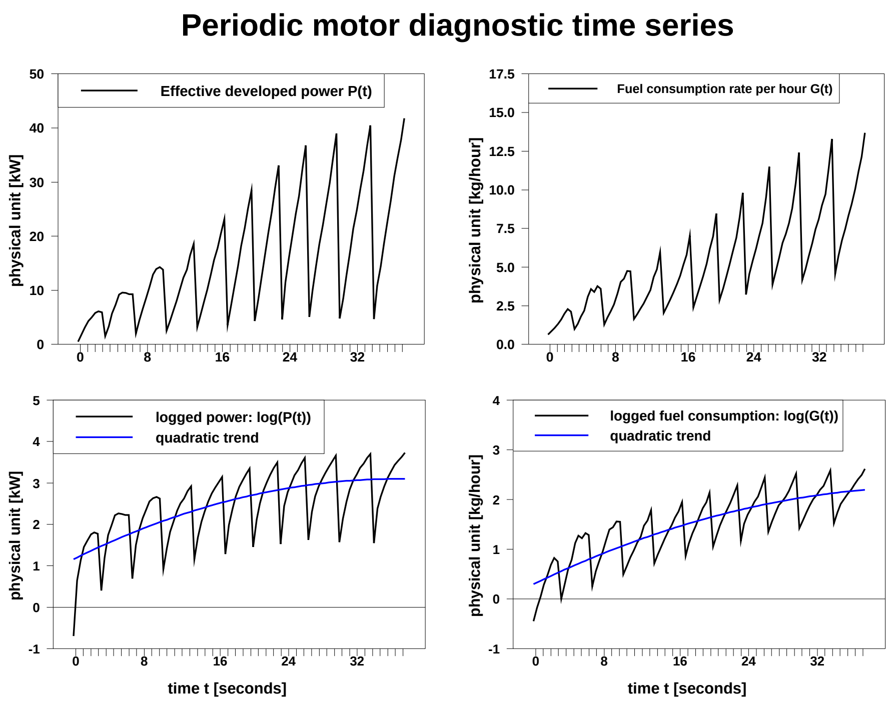

Figure 1: Initial and logged time series

2. Presentation of the Analysed Time Series

The most important *diagnostic* parameters, reflecting the assemblage quality of the engine, have been selected. There are $k = 2$ diagnostic parameters, describing the work of the internal-combustion engine: these are *effective power* and the *quantity of fuel consumed per unit of time*, here, the so-called *fuel consumption rate per hour*. Moreover, the fact that the *fuel consumption rate per hour* is inside of some limits tells about normal operating of the power supply system and ignition[1]. During the tests, the current information of $k = 2$ tested object states constitute the sequences of diagnostic parameters in function of time t, the *developed power* $P(t)$ in [kWatt] and the *quantity of fuel consumed per hour* $G(t)$ in [Kg/hour][2]. They are sampled through a measurement procedure with $\Delta t = 0.4$ sec over the period $[0; T]$, giving samples $P_j = P(j\Delta t)$, $G_j = G(j\Delta t)$, $j = 0, ..., N-1$. Thus, there are $N = 97$ equidistant samples within the period $[0; T]$, generating two time series, noted as $\{P_j; j = 0, ..N-1\}$, respectively $\{G_j; j = 0, ..N-1\}$ or also presented as vectors, $[P_0, ..., P_{N-1}]'$ respectively $[G_0, ..., G_{N-1}]'$. For notational purpose, it is stated that time series $\{P_j\}$, respectively $\{G_j\}$, presented with index $j = 0, ..., N-1$, are the collections of the samples $P_j = P(j\Delta t)$ of the *developed power*, respectively of the samples $G_j = G(j\Delta t)$ of the *quantity of fuel consumed per hour* at the moments $t = j\Delta t$.

The graphs of the both time series are presented (Fig. 1, upper part)[3].

[1]For the present investigation three time series, the *quantity of fuel consumed per hour*, the *torque* M [Nm] and the *engine speed* n, counted in *rotations per minute* [RPM] $[min^{-1}]$ are needed. The effective power P [kWatt] of engine is not measured, but it is calculated through the formula

$$P = M \cdot n \ \ [\frac{kWatt}{9550}]. \tag{1}$$

The factor 9550 is determined from unit calculations, when numerical values for n are entered in RPM and M in [nM]. The calculations of the physical units are as follows:

$$[P] = [Nm][\frac{2\pi}{60sec}] = [\frac{Nm}{sec}]\frac{2\pi \cdot 1000}{1000 \cdot 60} = [Watt]1000(\frac{2\pi}{1000 \cdot 60}) = 1\frac{kWatt}{9550}, \tag{2}$$

[2]http://www.thefreedictionary.com/fuel+consumption+rate

[3]For the computations, the software packages RATS and CATS is used, see Doan ([4]) and Hansen [14].

3. Transformation of Initial Time Series

The series *effective power* $\{P_j\}$ and the *fuel consumption rate per hour* $\{G_j\}$, see Fig. 1, exhibit a trend and increasing amplitudes in nearly exponential way. Neither harmonic analysis nor cointegration analysis can be realised for time series with such properties. Power spectral densities of such series show grave distortions. Linear trend presents itself at high frequencies and the increasing amplitudes give leakage also in the high frequencies, both covering the real frequencies and amplitudes and with differencing of exponentially growing time series stationarity cannot be attained, because its second moment does not become constant.

For this reason a transformation of the measured times series is necessary to eliminate the exponential growth. The initial time series are logged, eliminating the exponential growth. Clearly a non linear trend remains visible that is

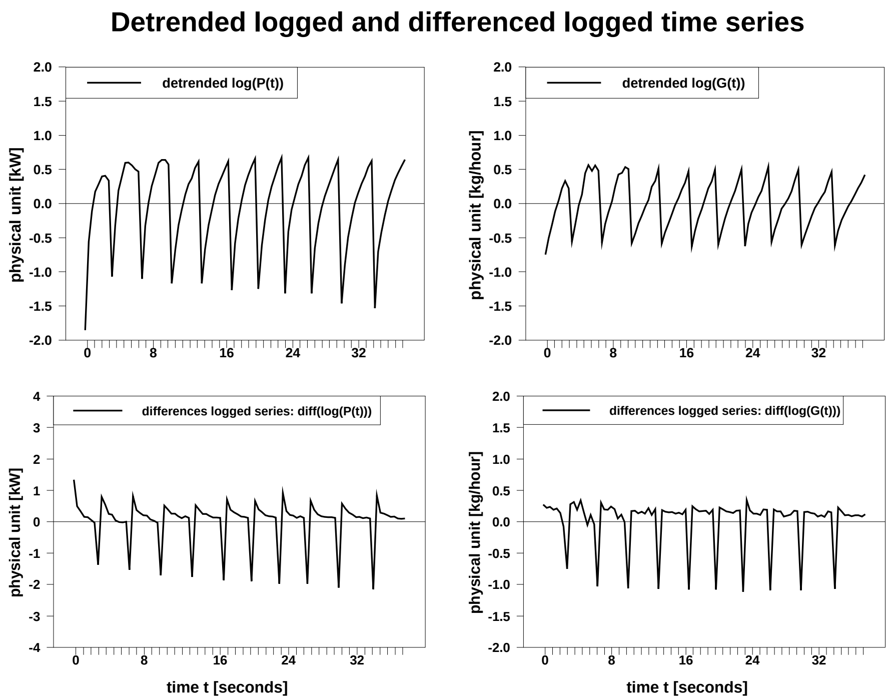

Figure 2: Logged initial time series without quadratic trend

proposed to be modeled by a quadratic trend. The coefficients are estimated through an OLS procedure[4]. The residuals of this operation are the detrended

[4]The RATS source routine LINREG is applied on the logged time series on which a

logged time series on which the harmonic analysis will be performed to discover essential frequencies and the noise, see Fig. 3.

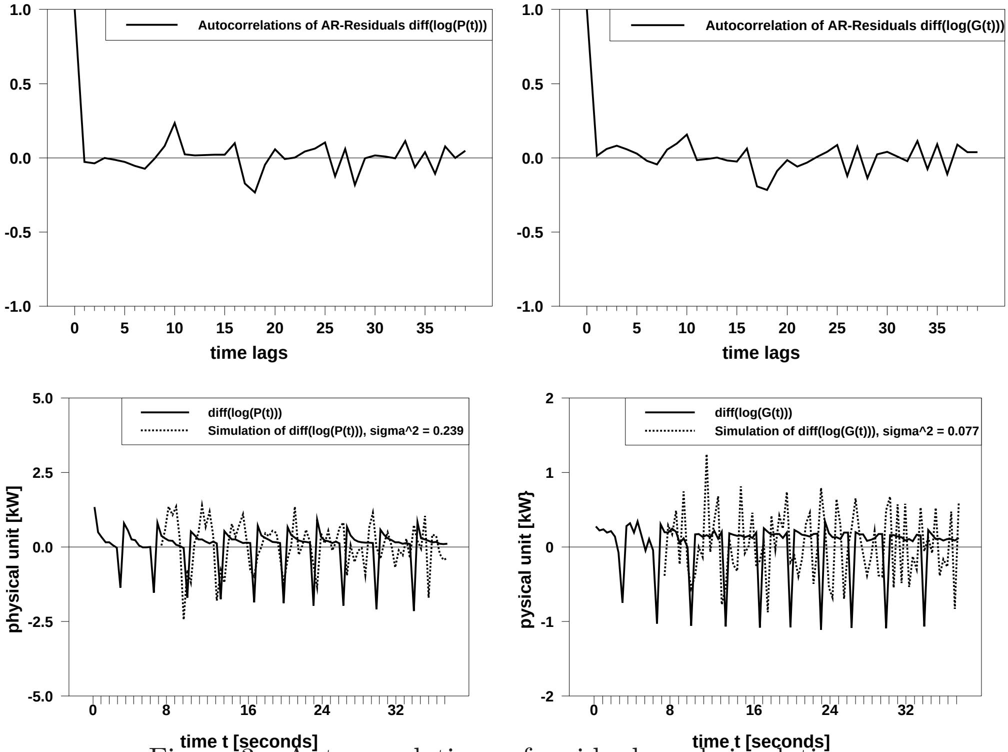

Figure 3: Autocorrelations of residuals and simulations

On the other hand the cointegration analysis will be directly applied on the logged time series, see Fig. 1, lower figures.

4. Cointegration Analysis

Looking at the graphs of the the series $\{P_j\}$ and $\{G_j\}$, respectively $\{log(P_j)\}$ and $\{log(G_j)\}$, see Fig. 1, cointegration cannot be excluded. The Johansen methodology [16] is used to investigate cointegration.

---

quadratic trend is applied. The residuals of this calculations are the detrended logged time series.

4.1. Definitions an Concepts

First, let's present a definition of *cointegration* in slightly modernized form.

Definition 1: A series with possible polynomial component which has a stationary, invertible, ARMA representation after differencing d times, is said to be integrated of order d, denoted $x_t \sim I(d)$[5]. *The components of the vector* x_t *are said to be cointegrated of order d, b, denoted* $x_t \sim CI(d,b)$, *if (i) all the components are* $I(d)$; *(ii) there exists a vector* $\beta(\neq 0)$ *so that* $z_t = \beta' x_t \sim I(d-b)$ $b > 0$. *The vector* β *is called the cointegration vector.*

Compared to the definition of Engle and Granger ([5], p. 252-253), it is admitted that the initial process has some *polynomial component*, because following Hamilton ([13], p. 435-437), we allow a $I(0)$ process to have possible nonzero mean, being also stationary, see also Hayashi ([15], p. 558). This *non zero mean* in $I(0)$ after d differencing operations results from a polynomial in the time variable t of order d.

It is immediately clear that cointegration necessitates $d > 0$. In the present case, we will deal with $d = 1$.

The initial time series exhibiting exponential growth, $\{P_j\}$ and $\{G_j\}$, see Fig. 1, cannot be integrated $I(d)$, for any order d, because with differencing, the variation remains exponential. By definition exponential variation of a time series gives necessarily non stationarity, see Hamilton ([13], p. 448, [15.5.1]). Taking the logarithms of the series $\{P_j\}$, $\{G_j\}$, one gets the series $\{log(P_j)\}$, $\{log(G_j)\}$, exhibiting no longer exponential growth, see Fig. 1. As it is a prerequisite of cointegraiton that the differenced series become stationary, cointegration analysis is proposed for the series $\{log(P_j)\}$ and $\{log(G_j)\}$. This will give a functional relationship between both time series $\{P_j\}$ and $\{G_j\}$.

We are specially interested in obtaining the linear cointegration relation

$$log(P_j) \cong b + a\, log(G_j) = b + log({G_j}^a); \quad j = 0, ..., N-1, \tag{3}$$

meaning a linear equation between both logged time series $\{log(P_j)\}$ and $\{log(G_j)\}$, where the coefficients a and b have to be estimated. By delogging (3) one gets with $A = e^b$ the functional relationship

$$P_j = A{G_j}^a \quad ; \quad j = 0, ..., N-1, \tag{4}$$

between the time series $\{P_j\}$ and $\{G_j\}$. Let's now move to the cointegration analysis.

[5] the chosen index is t as in the original literature, see Engle and Granger [5]

4.2. Cointegration of $log(P_j)$ and $log(G_j)$

It is visible that the periodicity of the series $\{log(P_j)\}$ and $\{log(G_j\})$ is not regular. Dummy seasonals are not appropriated to model such a periodicity, as preliminary investigations have revealed. For this reason the idea of the analysis is to grasp the periodicity by lagged values, also for logged and differenced series. For this purpose, the analysis starts to demonstrate that the series $\{log(P_j)\}$ and $\{log(G_j)\}$ are integrated $I(1)$, see *Definition* of Engle and Granger ([5], p. 252). Due to the above definition, we assume that $\{log(P_j)\}$ and $\{log(G_j)\}$ exhibit *unit roots* of order $d = 1$, the 1-differenced series $\{\Delta log(P_j)\}$ and $\{\Delta log(G_j)\}$ are $I(0)$, stationary with eventually a non zero mean and have an invertible ARMA representation.

Unit roots. First, the property of unit roots is investigated. The Dickey-Fuller unit-root tests have been realised. The first condition of property of

j (1)	series (2)	lags (3)	Dickey-Fuller (4)	c.v. (5%) (5)	trend (6)	unit-root (7)
1	$log(P_j)$	9	-2.58616	-2.894	no	yes
2	$\Delta log(P_j)$	10	-4.83209	-3.462	no	no
3	$log(G_j)$	9	-0.04147	-2.894	no	yes
4	$\Delta log(G_j)$	12	-5.54978	-3.462	no	no

Table 1: Dickey-Fuller unit-root tests

integration, the hypothesis of existence of unit roots for $\{log(P_j)\}$ and $\{log(G_j)\}$ cannot be rejected, see Table 1.

The stationary, invertible ARMA representation of the differenced series. Through a cascade of OLS F Wald tests it is shown that the components of the vector $\mathbf{x}_j = [x_{1j}, x_{2j}]' = [\Delta(log(P_j)), \Delta(log(G_j))]'$ of differenced series have a stationary representation

$$\Delta x_{ij} = \mu + \sum_{m=1}^{p} \delta_{im} \Delta x_{i(j-m)} + \varepsilon_{ij}\,;\, \varepsilon_{ij} \sim iid(0, \sigma_i^2), i = 1, 2. \qquad (5)$$

The starting order p of the $VAR(p)$ model has to be chosen. The visible periodicity is $p \geq 9$, see Fig. 1. For this reason, one sets $p = 13$ to surely grasp the whole visible periodicity stochastically, estimating in a first step a full $VAR(13)$ model. The reduction of parameters is conducted through a cascade of consecutive joint OLS F Wald tests, presented in Table 2.

The Table 2 col. (1) gives the name of the series, col. (2) gives the order p of the estimates AR model, col. (3) gives the estimated constant μ of the actual

series (1)	order p (2)	μ (3)	lags (4)	reduction of lags (5)	OLS F Wald (6)	Dickey-Fuller (7)	estimated variance (8)
$log(P_j)$	13	0.18083	1 to 13	9 to 13	0.5274	1.9367	0.2329
	8	0.15567	1 to 8			1.9395	*0.2398*
$log(G_j)$	13	0.13846	1 to 13	9 to 13	0.8715	1.9597	0.0725
	8	0.13845	1 to 8			1.8961	*0.0771*

Table 2: Dickey-Fuller unit-root tests

model, col. (4) indicates the lags in the actual model, col. (5) gives the lags to be eliminated in the next estimation, because the corresponding coefficients are only randomly different from zero, col. (6) gives the OLS F Wald statistics, as these values are comfortable greater than zero, it means that the null hypothesis of no reduction of the lags in col. (5) is not refused. In col. (7), the Dickey-Fuller statistics indicates a comfortable lack of serial correlation, the rules being: $1.75 < DW < 2.25$, col. (8) indicates the estimated variances of the AR models.

The estimated AR-models of both series are as follows:

$$\begin{aligned}\Delta log(P_j) = \underset{(0.054)}{0.156} - \underset{(0.105)}{0.799}\,\Delta log(P_{j-1}) - \underset{(0.107)}{0.762}\,\Delta log(P_{j-2}) - \underset{(0.108)}{0.734}\,\Delta log(P_{j-3}) \\ - \underset{(0.108)}{0.730}\,\Delta log(P_{j-4}) - \underset{(0.108)}{0.725}\,\Delta log(P_{j-5}) - \underset{(0.107)}{0.717}\,\Delta log(P_{j-6}) \\ - \underset{(0.105)}{0.726}\,\Delta log(P_{j-7}) - \underset{(0.100)}{0.376}\,\Delta log(P_{j-8});\quad \varepsilon_{jP} \sim iid(0, \hat{\sigma}_P^2 = 0.2398)\end{aligned} \tag{6}$$

$$\begin{aligned}\Delta log(G_j) = \underset{(0.033)}{0.138} - \underset{(0.106)}{0.730}\,\Delta log(G_{j-1}) - \underset{(0.106)}{0.723}\,\Delta log(G_{j-2}) - \underset{(0.107)}{0.686}\,\Delta log(G_{j-3}) \\ - \underset{(0.105)}{0.691}\,\Delta log(G_{j-4}) - \underset{(0.104)}{0.689}\,\Delta log(G_{j-5}) - \underset{(0.105)}{0.681}\,\Delta log(G_{j-6}) \\ - \underset{(0.103)}{0.696}\,\Delta log(G_{j-7}) - \underset{(0.104)}{0.362}\,\Delta log(G_{j-8});\quad \varepsilon_{jG} \sim iid(0, \hat{\sigma}_G^2 = 0.0771)\end{aligned} \tag{7}$$

The models (6), (7) are pure AR-models. Invertibility is out of question, because it only concerns MA-parts. On the other hand, the null hypothesis of a suitable AR representation of the differenced series $\{\Delta log(P_j)\}$ and $\{\Delta log(G_j)\}$ through (6), (7) cannot be rejected, due to the results of the OLS F Wald tests and DW tests, see Table 2.

Thus, the conditions of *integration* (stationarity and ARMA representation after differencing) are fulfilled for the logged series $\{log(P_j)\}$, $\{log(G_j)\}$ and cointegration can be investigated.

4.3. Setting up a VAR(p) Model

Analysing the definition of cointegration, it is evident that the multivariate vector autoregressive VAR models are the prerogatives of cointegration analysis.

Therefore, a vector autoregressive model of p lags without seasonals, designed as $VAR(p)$, with linear trend is set up, here with time variable t,

$$\mathbf{x}_t = \boldsymbol{\alpha} + \boldsymbol{\mu} t + \sum_{m=1}^{p} \mathbf{A}_m \mathbf{x}_{t-m} + \boldsymbol{\varepsilon}_t \;\; ; \;\; t = 0, ..., N-1, \tag{8}$$

where $\boldsymbol{\alpha}$ is the (2×1)-vector of intercept terms, $\boldsymbol{\mu}$ is the trend (2×1)-vector, $\mathbf{A}_m$ are (2×2)-matrices of coefficients and $\boldsymbol{\varepsilon}_t$ is a (2×1)-vector of error terms. For parameter estimation, the least squares method is applied, the order of the model p is determined with the help of the Akaike Information Criterion (AIC)[6].

The model (8) represents a vector autoregressive model $VAR(p)$ that can be transformed into the error-correction form. Thus, the level $\boldsymbol{\alpha}$ disappears, the trend gradient $\boldsymbol{\mu}$ becomes the level. The corresponding algebra is presented in Hamilton, see ([13], p. 517). One gets

$$\Delta \mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\Pi} \mathbf{x}_{t-1} + \sum_{m=1}^{p-1} \boldsymbol{\Pi}_m \Delta \mathbf{x}_{t-m} + \boldsymbol{\varepsilon}_t \;\; ; \;\; t = 0, ...N-1, \tag{9}$$

with matrix $\boldsymbol{\Pi} = \mathbf{I} - \mathbf{A}_1 - \mathbf{A}_2 - ... - \mathbf{A}_p$. The rank of $\boldsymbol{\Pi}$ is equal to the number of independent cointegrating vectors. In the present case, one is interested for $rank(\boldsymbol{\Pi}) = 1$, where there is one cointegrating vector[7].

4.4. The Johansen Methodology to Calculate Cointegration Relations

A VAR model in error-correction form is set for the series $log(P_j)$ and $log(G_j)$. The optimal number of p lags of the corresponding $VAR(p)$ is determined with the AICC and SBC/BIC criterion[8]. The AIC indicates lag $p = 8$ and the SBC/BIC indicates lag $p = 1$. We decide to take the average $p = 5$.

Now the *Johansen cointegration rank test* statistics are summarised [18, 19]. There is the maximum eigenvalue and the trace test, whose critical values have been tabulated first by Johansen and Juselius [20, 21].

---

[6]The RATS-routine @*selectvarlag*, written by Tom Doan in June 2005, see also RATS 7 manual [4], is used to calculate the AIC criterion values.

[7]The number of distinct cointegrating vectors is obtained by calculating the number of significant eigenvalues λ of equation $|\lambda S_{kk} - S_{k0} S_{00}^{-1} S_{0k}| = 0$. The matrices $S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R'_{jt}, \; ; \; i,j = 0,..,k$ are product moment matrices of residuals R_{it}. For the residuals R_{it} see Johansen [17, 18].

[8]The AIC and SBC/BIC criterion is calculated with the RATS-routine @varlagselect.src, see Doan [4].

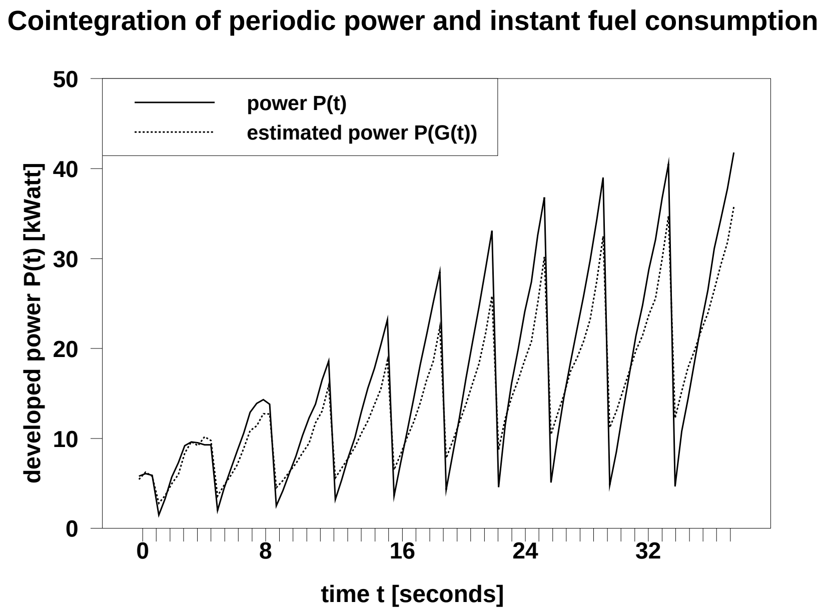

Figure 4: Cointegration relation between P ($kWatt$) and G ($kg/hour$)

The trace statistics. If the null hypothesis is that the matrix $\mathbf{\Pi}$ (9) has $rank(\mathbf{\Pi}) = r \leq k$ against the alternative hypothesis $rank(\mathbf{\Pi}) > k$. Then use the λ_{trace} statistics, see Enders [7].

$$\lambda_{trace} = -T \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i) \tag{10}$$

The maximum eigenvalue statistics. If the null hypothesis is that the matrix $\mathbf{\Pi}$ (9) has $rank(\mathbf{\Pi}) = r$ against the alternative hypothesis $rank(\mathbf{\Pi}) = r + 1$, then use the λ_{max} *statistics*, see Enders [6].

$$\lambda_{max} = -T \ln(1 - \hat{\lambda}_{r+1}) \tag{11}$$

The λ_{max} and the λ_{trace} test are used together and the standard tables produced by CATS will be presented later, the interpretation of the results are straight forward.

Procedure of the Johansen cointegration rank test. CATS calculates all the eigenvalues, the λ_{max} and the λ_{trace} statistics and presents them together with the critical values $\lambda_{max(1-\alpha)}$ and $\lambda_{trace(1-\alpha)}$ for the significance level set here at $\alpha = 0.1$. The eigenvalues appear in decreasing order and the hypothetical rank value r in ascending order.

A constant restricted to the cointegration space. Due to the AICC/ SCB/ BIC criterion, the lag $p = 5$ is set for the VAR model. A $VAR(5)$ model is calculated[9]. The λ_{max} and λ_{trace} statistics are displayed along with the critical values of Johansen and Nielsen [19] for $\alpha = 0.1$ in Table 3. A constant has been restricted to the cointegration space, see Hansen, ([14], p. 14), because one wants the cointegration relation to contain an additive constant b as in (3).

$$\widehat{log(P_j)}(log(G_j)) = b + a\ log(G_j) \quad ; \quad j = 0, ..., N-1 \tag{12}$$

The rank tests indicate that the eigenvalue $\lambda_1 = 0.2532$ is significant[10], whereas

eigenvalues (1)	λ_{max} (2)	λ_{trace} (3)	H_0 (4)	H_1 (5)	λ_{max90} (6)	$\lambda_{trace90}$ (7)
$\lambda_1 = 0.2532$	26.86	32.47	$r = 0$	$p - r = 2$	10.29	17.79
$\lambda_2 = 0.0592$	5.62	5.62	$r = 1$	$p - r = 1$	7.50	7.50

Table 3: Johansen cointegration rank tests for $\{log(P_j)\}$ and $\{Log(G_j)\}$

the second eigenvalue $\lambda_2 = 0.0592$ is not significant[11]. The cointegration rank tests indicate rank $r = 1$. Then, the absolute values of both first of the $kp = 2 \times 5 = 10$ eigenvalues of the 'companion matrix' are computed, see Juselius ([22], p. 50). There is $u = 1$ eigenvalue equal 1.000. The second is 0.6464, too 'far away' from the border of the unit circle, indicating the rank $r = k - u = 2 - 1 = 1$. This means that the Johansen cointegration rank tests and the 'companion matrix' indicate cointegration rank $r = 1$.

There is no longer serial correlation in the residuals, as the LM test of Breusch-Godfrey [10], [2] shows. Indeed, one finds: $LM(4)$, $P(\chi^2(4) \geq 3.681) = 0.45$. Then, the distributions of the residuals are far from normal. But with

[9]The AIC/SCB/BIC criterion are calculated for every model with the RATS-routine @varlagselect.src, see Doan [4]

[10]The statistics $\lambda_{max} = 26.86 > \lambda_{max(1-\alpha)} = 10.29$, means that the null hypothesis $r = 0$ is rejected against the alternative hypothesis $r = 1$. Then, the statistics $\lambda_{trace} = 32.47 > \lambda_{trace(1-\alpha)} = 17.79$, means that the null hypothesis $r = 0$ is rejected against the alternative hypothesis $r > 0$. Thus, both rank tests postulate $r = 1$.

[11]The statistics $\lambda_{max} = 5.62 < \lambda_{max(1-\alpha)} = 7.50$, means that the null hypothesis $r = 1$ is not rejected against the alternative hypothesis $r = 2$. Then, the statistics $\lambda_{trace} = 5.52 < \lambda_{trace(1-\alpha)} = 7.50$, means that the null hypothesis $rleq1$ is not rejected against the alternative hypothesis $r > 1$. Thus, both rank tests postulate $r = 1$

these test results, *one* cointegration relation can be justified and calculated

$$\widehat{log(P_j)}(log(G_j)) = 1.028 + 0.974 log(G_j) \quad ; \quad t = 0, ..., (N-1). \tag{13}$$

Equation (13) is delogged, giving

$$\widehat{P_j(G_j)} = 2.80\ G_j^{0.974} \quad ; \quad t = 0, ..., (N-1). \tag{14}$$

The cointegration relation (14) is graphed in Fig. 4.

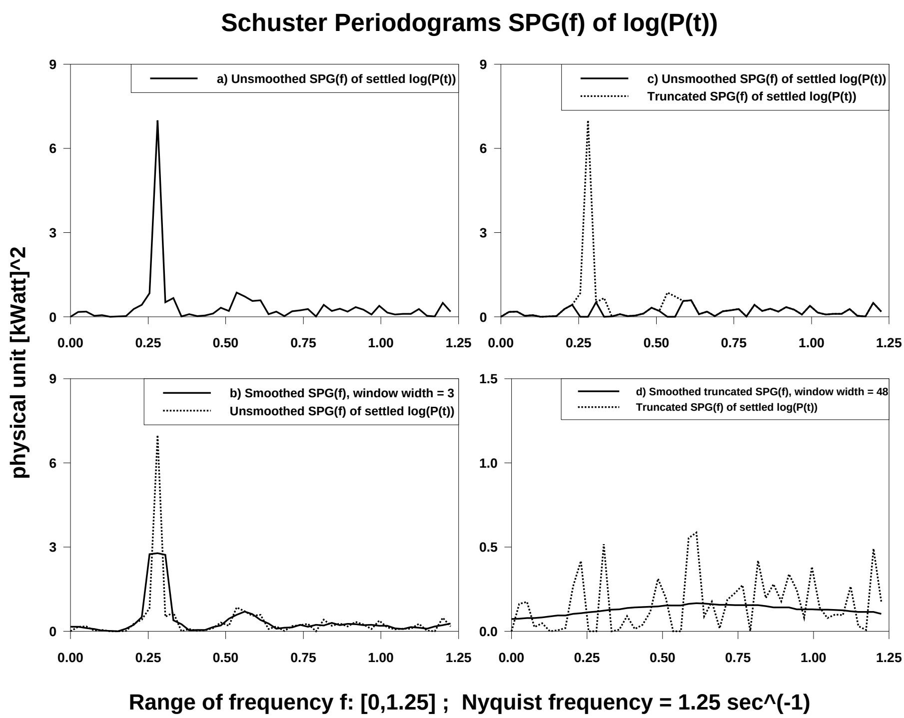

Figure 5: Periodogram of settled $log(P(t))$

5. Harmonic Analysis

The aim of spectral or harmonic analysis is to detect existing significant *frequencies* and *amplitudes* within a signal, eventually affected by *noise*, characterised typically by a normal distribution. As the measured *effective power* and the *fuel consumption rate per hour* are discrete sampled time series P_j and G_j, it follows from the theory of Fourier Transforms that their spectrum is also discrete, see Priestley ([26], p. 390). There are $k = 2$ time series. The usual model applied to data with discrete spectrum to analyse essential fluctuations is given by the *harmonic model*, including a *white noise* error term $\varepsilon_{ij} \sim iid(0, \sigma_i^2)$. Let's set $\mathbf{x}_j = [x_{1j}, x_{2j}]' = [log(P_j), log(G_j)]'$,

$$x_{ij} = \sum_{m=1}^{K} a_{im} cos(2\pi\nu_{im} j\Delta t + \phi_i) + \varepsilon_{ij} \quad ; \quad j = 0, 1, ..., (N-1), \ \ i = 1, 2 \quad (15)$$

where N is the known sample size. The number K gives the number of peaks with harmonics, a_m are the essential amplitudes and σ_i^2 is the noise variance. The phases ϕ_i are not of special interest, because they represent only a lateral move. The harmonic components $x_{ij} = x_i(j\Delta t)$ give the value at moment $j\Delta t$.

5.1. Definitions

The Nyquist-Shannon sampling theorem states that perfect reconstruction of a signal is possible when the sampling frequency is greater than twice the maximum frequency of the signal being sampled, or equivalently, when the Nyquist frequency (half the sample rate) exceeds the highest frequency of the signal being sampled. If lower sampling rates are used, the original signal's information may not be completely recoverable from the sampled signal [2]. For example, if a signal has an upper band limit of 100 Hz, a sampling frequency greater than 200 Hz will avoid aliasing and would theoretically allow perfect reconstruction[12].

Consider *one* complex stationary random process $x(i\Delta t) := x_i$, $i = 0, ..., (N-1)$, sampled at a sampling interval Δt, the time variable is here i. This means, there are then N contiguous measures x_i, $i = 0, ..., N-1$. Then, x_i^* meaning the complex conjugate of x_i. The sampling interval between two measurement is Δt.

The *power spectral density* (PSD), which is denoted by $P_{xx}(f)$, the Nyquist frequency being $Ny = \frac{1}{2\Delta t}$ of x_i is defined with the imaginary root $j = \sqrt{-1}$ as

$$P_{xx}(f) = \sum_{k=-\infty}^{\infty} r_{xx}[k] \, exp(-j2\pi f k\Delta t) \quad ; \quad -Ny \leq f \leq Ny. \quad (16)$$

where $r_{xx}[k]$ is the autocorrelation function (ACF) of x_i defined with the expectation operator $\mathcal{E}$ as

$$r_{xx}[k] = \mathcal{E}(x_i^* x_{i+k}) \quad (17)$$

[12] http://en.wikipedia.org/wiki/Sampling_ rate,18.11.2012

The spectral estimation problem can be summarized as follows: Based on the N contiguous observations x_i, $i = 0, ..., (N-1)$ of a single equidistant realisation with sampling interval Δt of a stationary random process, it is wanted to estimate the PDS within the the frequency interval $[Ny \leq f \leq Ny]$, see Kay ([23], p. 4).

It can be shown that the PSD (16) can equivalently be defined as

$$P_{xx}(f) = \lim_{M \to \infty} \mathcal{E}\{\frac{1}{2M+1} \mid \sum_{i=-M}^{M} x_i \, exp(-j2\pi f i \Delta t) \mid^2\}. \quad (18)$$

Neglecting the expectation operator $\mathcal{E}$, considering that the available data are

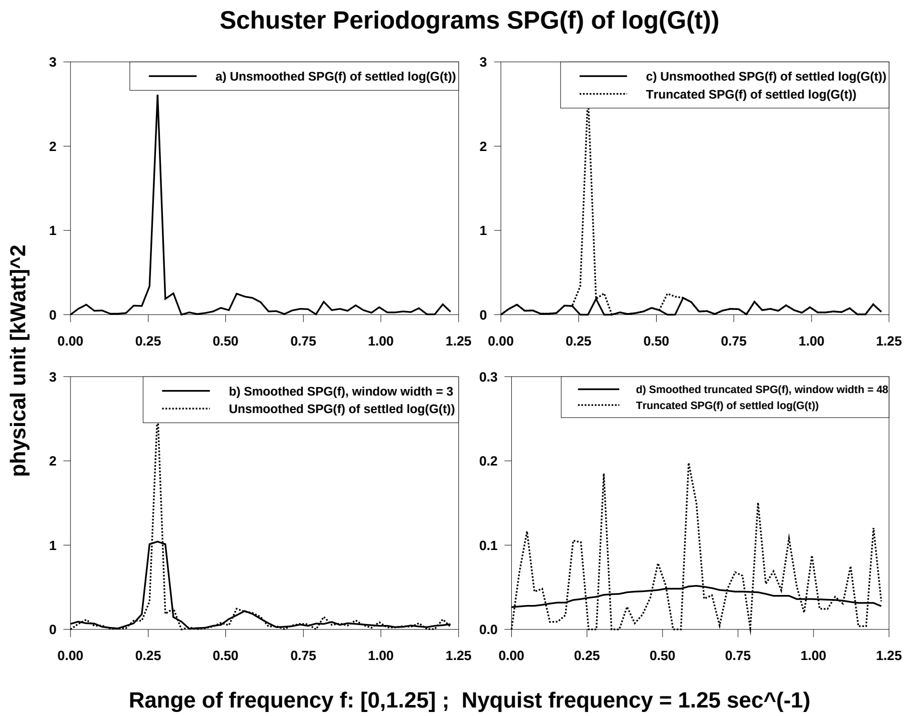

Figure 6: Periodogram of settled *log(G(t))*

$x_0, x_1, ..., x_{N-1}$, the frequency f is evaluated as usual at equidistant frequencies $f = f_k = \frac{k}{N\Delta t}$ over $k = 0, ..., \frac{N-1}{2} - 1$ (here $N = 97$ uneven) because of symmetry properties, the *power spectral density* (PSD) is estimated by the *periodogram spectral estimator* of Schuster [29], the so-called *Schuster periodogram* (SPG) defined with the *Discrete Fourier Transform*, where the time interval Δt

can be reduced in the fraction $-\frac{j2\pi ki\Delta t}{N\Delta t}$ of the exponent of the function $exp(.)$

$$\hat{P}_{PER}(f_k = \frac{k}{N\Delta t}) = \frac{1}{N} \mid \sum_{i=0}^{N-1} x_i \; exp(-\frac{j2\pi ki}{N}) \mid^2 \quad ,k = 0, ..., 47. \tag{19}$$

The $SPG(f)$ of the settled $log(P(t))$ and $log(G(t))$ are shown in Fig 5,6 a). Then they are windowed with a flat window $w(d)_s$ of $width = d$ defined by the discrete convolution

$$\hat{P}_{WIN}(f_k) = \sum_{s=-\infty}^{\infty} w(d)_s \; \hat{P}_{PER}(f_{k-s}) \quad ; \quad k = 0, ..., 47. \tag{20}$$

The windowed $SPG(f)$ with $w(3)_s$ are presented in Fig. 5,6 b) for the settled $log(P(t))$ and $log(G(t))$. Both $SPG(f)$ show maximal peaks at the period $P = 3.564sec \sim 3.6sec$. Then, for the $SPG(f)$ of the settled $log(P(t))$, the peaks $SPG(f) \geq 0.65$ are replaced by 0, for the $SPG(f)$ of the settled $log(G(t))$, the peaks $SPG(f) \geq 0.2$ are replaced by 0 as presented in Table 4, observing that $K = 2^{13}$. The truncated $SPG(f)$ are presented in Fig. 5,6 c). The windowed truncated $SPG(f)$ with $w(48)_s$ are presented in Fig. 5,6 d). The mean of the windowed truncated $SPG(f)$ is calculated. This is an estimation of the *variances* of the error terms of the harmonic models (15). This means: for the settled $log(P_j)$, the estimation of the variances of the error term ε_{jP} is $\hat{\sigma}_P^2 = 0.12923$, for the settled $log(G_j)$ the estimation of the variance of the error term ε_{jG} is $\hat{\sigma}_G^2 = 0.03926$.

frequency f $[Hz]$ (1)	period $P = \frac{1}{f}$ [sec] (2)	unsmoothed $SPG(f)$ of $log(P(t))$ (3)	truncated $SPG(f)$ of $log(P(t))$ (4)	unsmoothed $SPG(f)$ of $log(G(t))$ (5)	truncated $SPG(f)$ of $log(G(t))$ (6)
0.255102	3.920000	0.834049	0	0.335264	0
0.280612	**3.563639**	6.996398	0	2.606280	0
0.331633	3.015381	0.668490	0	0.251441	0
0.535714	1.866668	0.859054	0	0.246101	0
0.561224	**1.781820**	0.717589	0	0.214363	0

Table 4: Truncation of Schuster periodograms

[13] There exists the Fisher $g-$statistics, determining *essential amplitudes* of a harmonic model (15), comparing the essential peaks of a SPG with the sums of all the peaks, see Fisher, [8]. In the present case, the truncation value (0.65, respectively 0.20) are chosen 'by eye', in analogy with the idea of that test, selecting the main huge amplitude at period $p \sim 3.6sec$ and one harmonic at period $p \sim 1.8sec$

6. Conclusions

In this study a delogged cointegration relation (14) between the settled logged time series of two parameters of *carburetor internal combustion engines* during industrial testing, the *developed power* and the *fuel consumption rate per hour* could be established. Second, as the data show strong periodicity, an appropriated *harmonic analysis* is performed, the frequency of the period of $p \sim 3.6$ sec with one harmonic is clearly visible in the *Schuster periodograms* (SPG), an estimation of the *power spectral densities* (PSD). The estimation of the *variances* of the noise terms of the underlying harmonic models (15) could be established through a procedure of separation of the noise part and the part of essential amplitudes in the *Schuster periodogram*.

Acknowledgments

The authors express their deep gratitude to the Department of Quantitative Economics of the University of Fribourg for the encouragement and support of this research, granted by the SNSF-SCOPES project Nr. 127 962.

References

[1] G.E. Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis, Forecasting and Control*, Prentice Hall (1994).

[2] T.S. Breusch, Testing for autocorrelation in dynamic linear models, *Australian Economic Papers*, **17** (1979), 334-355.

[3] P.J. Brockwell, R.A. Davis, *Time Series: Theory and Methods*, 2-nd Ed., Springer-Verlag (1991).

[4] T.A. Doan, *RATS Software Package, User's Manual*, Version 8.0., Estima, Evanston, Illinois (August 2011).

[5] R.E. Engle, C.W.J. Granger, Cointegration and error correction: Representation, estimation and testing, *Econometrica*, **55** (1987), 251-276.

[6] W. Enders, *Applied Econometric Time Series*, New York, John Wiley and Sons (1995).

[7] W. Enders, *RATS Handbook of Econometric Time Series*, John Wiley and Sons, Inc. New York (1996).

[8] R.A. Fisher, Tests of significance in harmonic analysis, *Proc. Roy. Soc. Ser. A*, **125** (1929), 54-59.

[9] P.H. Franses, *Periodicity and Stochastic Trends in Economic Time Series*, Oxford, University Press (1996).

[10] L.G. Godfrey, Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, *Econometrica*, **46** (1978), 1293-1302.

[11] V. Golikova, E. Pervukhina, P. Sopin, *Statistical Modeling of Machines on Diagnostic Parameters Problems of Machine Building and Reliability*, Russian Academy of Sciences, Moscow, Russia, **6** (2008), 89-95.

[12] V. Golikova, E. Pervukhina, J.-F. Emmenegger, Cointegration of control parameter time series measured on motors at the stage of assemblage, *International Journal of Pure and Applied Mathematics*, **42**, No. 1 (2008), 139-151.

[13] J.D. Hamilton, *Time Series Analysis*, Princeton, University Press (1994).

[14] H. Hansen, *CATS in RATS: Cointegration Analysis of Time Series*, Handbook for the Software Package CATS, Illinois, ESTIMA (1995).

[15] F. Hayashi, *Econometrics*, Princeton University Press, Princeton and Oxford (2000).

[16] S. Johansen, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford, University Press (1995).

[17] S. Johansen, Statistical analysis of cointegration vectors, *Journal of Econometric Dynamics and Control*, Elsevier Science Publishers B.V., North-Holland, **12**, 231-254 (1988).

[18] S. Johansen, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, **59**, No. 6 (1991), 1551-1580.

[19] S. Johansen, B. Nielsen, *Asymptotics for Cointegration Rank Tests in the Presence of Intervention Dummies*, University of Kopenhagen, Institute of Mathematical Statistics, Denmark (1993).

[20] S. Johansen, K. Juselius, Testing structural hypotheses in a multivariate cointegration analysis of PPP and the UIP for UK, *Journal of Econometrics*, **53** (1992), 214-244.

[21] S. Johansen, K. Juselius, Maximum likelihood estimation and inference on cointegration with application to the demand of money, *Oxford Bulletin of Economics and Statistics*, **52** (1990), 169-209.

[22] K. Juselius, *The Cointegrated VAR Model. Methodology and Applications*, Oxford University Press Inc, New York (2006).

[23] Kay, Steven, M., *Modern Spectral Estimation, Theory and Applications*, PTR Prentice Hall Signal Processing Series, Englewood Cliffs, New Jersey (1987).

[24] *Technical Means for Diagnostics: Reference Book*, V. Klyuev, Editor. Moscow, Mashinostroenie (1989), In Russian.

[25] S.G. Makridakis, V. Wheelwright, E. McGee, *Forecasting: Methods and Applications*, New York, Wiley (1983).

[26] M.B. Priestley, *Spectral Analysis and Time Series*, Probability and Mathematical Statistics, Academic Press, Harcourt Brace Jovanovich, Publishers, London (1981).

[27] E. Pervukhina, V. Golikova, *Analysis of Non-stationary Stochastic Processes in Problems on Automation of Industrial Machine Testing, Assembling in Mechanical Engineering*, Instrument-Making, Moscow, Russia, No. 8 (2007), In Russian.

[28] V. Rybalko, *Parametric Diagnosing of Energetic Objects on the Base of Factor Analusis in Statistica Environment*, Exponenta Pro (2004), 78-83, In Russian.

[29] A. Schuster, On the investigations of hidden periodicities with application to a supposed 26-day period of meterological phenomena, *Terr. Mag. Atmos. Elect.*, **3** (1898), 13-41.

[30] V. Skurikhin, V. Kvachev, Y. Walkman, Y. Yakovenko, *Information Technologies in the Testing of Complex Objects*, Cityplace Kiev, Naukova Dumka (1990), In Russian.